NASA-CR-202426

Final Report covering the period of
6/1/1991 - 5/31/1993

Management of Knowledge Representation Standards Activities
NASA-Ames Contract NCC 2-719

Principal Investigator

Ramesh S. Patil
USC Information Sciences Institute
4676 Admiralty Way
Marina del Rey, CA 90292
Phone: (310) 822-1511
ramesh@isi.edu

Final Report covering the period of 6/1/1991 - 5/31/1993

Management of Knowledge Representation Standards Activities: The DARPA Knowledge Sharing Effort

Ever since the mid-seventies, researchers have recognized that capturing knowledge is the key to building large and powerful AI systems. In the years since, we have also found that representing knowledge is difficult and time consuming. In spite of the tools developed to help with knowledge acquisition, knowledge base construction remains one of the major costs in building an AI system: For almost every system we build, a new knowledge base must be constructed from scratch. As a result, most systems remain small to medium in size. Even if we build several systems within a general area, such as medicine or electronics diagnosis, significant portions of the domain must be represented for every system we create.

The cost of this duplication of effort has been high and will become prohibitive as we attempt to build larger and larger systems. To overcome this barrier we must find ways of preserving existing knowledge bases and of sharing, reusing, and building on them.

This report describes the efforts undertaken over the last two years under the NASA grant NCC 2-719 to identify the issues underlying the current difficulties in sharing and reuse, and a community wide initiative to overcome them. First, we discuss four bottlenecks to sharing and reuse, present a vision of a future in which these bottlenecks have been ameliorated, and describe the efforts of the initiative's four working groups to address these bottlenecks. We then address the supporting technology and infrastructure that is critical to enabling the vision of the future. Finally, we consider topics of longer-range interest by reviewing some of the research issues raised by our vision.

Sharing and Reuse of Knowledge

There are many senses in which the work that went into creating a knowledge-based system can be shared and reused. Rather than mandating one particular sense, the approach we have taken in this project seeks to support several of them. One mode is to reuse is through the exchange of techniques and detailed analysis of a domain. That is, the content of some knowledge-base or an implemented algorithm is not directly used, but the approach behind it is communicated in a manner that facilitates its reimplementation. Another mode of reuse is through the inclusion of an existing knowledge-base into a new system. That is, the content of some module is copied into another at design time and merged (possibly after extension or revision) into the new system. A third mode is through the run-time communication of knowledge, data or services. That is, one module invokes another either as a procedure from a function library or as a collaborated agent (specialist) in the problem-solving activity.

These modes of reuse do not work particularly smoothly today. Explaining how to reproduce a technique often requires communicating subtle issues that are more easily expressed formally; whether stated formally or in natural language, the explanations require shared understanding of the intended interpretations of terms. The reuse of a knowledge-base is only feasible to the extent that their model of the world is compatible with the intended new use. The reuse of external agents' knowledge or service is feasible only to the extent that we understand what requests the agent are prepared to accept. These difficulties in sharing and reuse and possible approaches to resolving them were studied at a 3 day workshop involving 40 top scientists in artificial intelligence, organized by ISI (and supported by DARPA, NSF, and industry).

Technical analyses of knowledge representation technology indicated four key impediments and four complementary areas in which development of common, agreed-upon conventions would enhance leverage between individual research efforts. The four areas of impediments are: (1) heterogeneous representations, (2) multiple dialects within language families, (3) lack of common conventions for communication among intelligent agents, and (4) model "mismatch" at the knowledge level. The proposed approaches to addressing each of these problem are: (1) mechanisms for translation between knowledge bases represented in different languages; (2) common versions of languages and reasoning modules within families of representational paradigm; (3) protocols for communication between separate knowledge-based modules, as well as between knowledge-based systems and databases; and, (4) libraries of ``ontologies," i.e., pre-fabricated foundations for application-specific knowledge bases in a particular topic area.

To further develop and refine the solutions proposed, develop conventions, and to test the solutions in implemented systems, Working groups (comprised of researchers from the DARPA AI community and other academic and industry volunteers) have been established for each of these four areas, and are developing draft specifications which are circulated for review both within the working groups and among the project participants at ISI. The specifications are fed back to various collaborating DARPA projects which are building relevant technology, so that their work can move toward providing preliminary implementations of the specifications. Under the current grant, ISI is responsible for supporting and coordinating the activities of these groups, and disseminating the result of these activities to the broader research community and potential users.

In the following sections we describe the impediments and the results of the working groups.

The impediments to knowledge sharing

Impediment 1. Heterogeneous Representations: There are a wide variety of approaches to knowledge representation, and knowledge that is expressed in one formalism cannot directly be incorporated into another formalism. However, this diversity is inevitable—the choice of one form of knowledge representation over another can have a big impact on a system's performance. There is no single knowledge representation that is best for all problems, nor is there likely to be one. Thus, in many cases, sharing and reusing knowledge will involve translating from one representation to another. Currently, the only way to do this translating is by manually recoding knowledge from one representation to another. We need tools that can help automate the translation process.

Impediment 2. Dialects within Language Families: Even within a single family of knowledge representation formalisms (for example, the KL-One family), it can be difficult to share knowledge across systems if the knowledge has been encoded in different dialects. Some of the differences between dialects are substantive, but many involve arbitrary and inconsequential differences in syntax and semantics. All such differences, substantive or trivial, impede sharing. It is important to eliminate unnecessary differences at this level.

Impediment 3. Lack of Communication Conventions: Knowledge sharing does not necessarily require a merger of knowledge bases. If separate systems can communicate with one another, they can benefit from each other's knowledge without sharing a common knowledge base. Unfortunately, this approach is not generally feasible for today's systems because we lack an agreed-on protocol specifying how systems are to query each other and in what form answers are to be delivered. Similarly, we lack standard protocols that would provide interoperability between knowledge representation systems and other, conventional software, such as database management systems.

Impediment 4. Model Mismatches at the Knowledge Level: Finally, even if the language-level problems previously described are resolved, it can still be difficult to combine two knowledge bases or establish effective communications between them. These remaining barriers arise when different primitive terms are used to organize them; that is, if they lack shared vocabulary and domain terminology. For example, the type hierarchy of one knowledge base might split the concept Object into Physical-Object and Abstract-Object, but another might decompose Object into Decomposable-Object, Nondecomposable-Object, Conscious-Being, and Non-Conscious-Thing. The absence of knowledge about the relationship between the two sets of terms makes it difficult to reconcile them. Sometimes these differences reflect differences in the intended purposes of the knowledge bases. At other times, these differences are just arbitrary (for example, different knowledge bases use Isa, Isa-kind-of, Subsumes, AKO, or Parent relations, although their real intent is the same). If we could develop shared sets of explicitly defined terminology, sometimes called ontologies, we could begin to remove some of the arbitrary differences at the knowledge level. Furthermore, shared ontologies could provide a basis for packaging knowledge modules—describing the contents or services that are offered and their ontological commitments in a composable, reusable form.

The Knowledge-Sharing Effort

The desire to collaborate through knowledge sharing and reuse has arisen within a segment of the broad knowledge representation community that is interested in scaling up to larger systems and that views the sharing and reuse of knowledge bases as a means to this end. Closely related to this effort is a concern for building embedded systems in which knowledge representation systems support certain functions rather than act as ends in themselves.

In particular, our goal is to support researchers in areas requiring systems bigger than a single person can build. These areas include engineering and design domains, logistics and planning domains, and various integrated modality areas (for example, multimedia interfaces). Researchers working on such topics need large knowledge bases that model complex objects; because these models drive complex systems, they cannot be skeletons. Putting together much larger systems, of which various stand-alone systems being built today are just components, is an interesting challenge.

The creation of such knowledge resources requires community wide effort. This effort engenders a need for agreed-on conventions to enable us to build pieces that fit together. Eventually, in pursuing the goal of large, shared knowledge bases as part of a nationwide information infrastructure, these conventions might become objects of study for the definition of more formal standards. Currently, however, the conventions are intended to support experiments in knowledge sharing among interested parties.

In the next section describes the activities of our four working groups on these foundation-laying activities. For each group, we summarize the problem being addressed, the approach being taken, and the outcomes sought.

Interlingua

The Interlingua Working Group is headed by Richard Fikes and Mike Genesereth, both of Stanford University.

Problem Formulation. The Interlingua Working Group focuses on the problems posed by the heterogeneity of knowledge representation languages. Specifically, to interchange knowledge among disparate programs (written by different programmers, at different times, in different languages), effective means need to be developed for translating knowledge bases from one specialized representation language into another. The goal of this group is to specify a language for communicating knowledge between computer programs (as opposed to a language for the internal representation of knowledge within computer programs). This language needs

(1) an agreed-on declarative semantics that is independent of any given interpreter,

(2) sufficient expressive power to represent the declarative knowledge contained in typical application system knowledge bases, and

(3) structure that enables semiautomatic translation into and out of typical representation languages.

Approach. This group is specifying a language (KIF [knowledge interchange format]) that is a form of predicate calculus extended to include facilities for defining terms, representing knowledge about knowledge, reifying functions and relations, specifying sets, and encoding commonly used nonmonotonic reasoning policies. The group is also conducting knowledge-interchange experiments to substantially test the viability and adequacy of the language. The experiments focus on developing and testing a methodology for semiautomatic translation to and from typical representation languages and the use of the interchange format as an intermodule communication language to facilitate interoperability.

Outcomes. The specification for interlingua will evolve in a set of layers. The innermost layer will be a core, analogous to the primitives in Lisp, providing basic representational and language extension functions. The next layer will provide idioms and extensions that make the language more usable, analogous to the set of functions provided by Common Lisp. This working group will be responsible for developing these specifications. Its output will be (1) a living document containing the current KIF specification, describing open issues, and presenting current proposals for modification and (2) a corpus of documented microexamples, using fragments of knowledge bases to illustrate how they translate into KIF and to point out open issues.